SYSTEM AND METHODOLOGY FOR THE STORAGE AND MANIPULATION OF DOCUMENTS

5

BACKGROUND OF THE PRESENT INVENTION

10 ## Field of the Invention

The present invention relates generally to the manipulation of stored

data, more particularly to systems and methodologies for the capture,

transmission, management, storage, retrieval and display of document images

in a shared-system environment using the Internet or other network.

Background of the Present Invention

Since the introduction of paper, there has been the problem of storing

documents and making them readily available for later use. As society entered

the Information Age, an ever growing mountain of paper documents became

5   increasingly difficult to store and manage. Certain document-intensive

industries, such as banking, have come under increasing pressure to manage

this problem.

With the advent of the computer and increasing data storage capabilities,

text and image-based data are now being electronically stored at an even greater

10   pace. Since geographical images require considerably more storage space and

processing power to manipulate than a simpler text-based system, conventional

commercial computer systems heretofore have been unable to adequately

service this growing segment of the industry, e.g., due to inadequate storage

capacities on other technological bottleneck. An additional problem with

15   image-based information is the inability at present to search the graphical image

itself and the need to correlate the image with sufficient relevant text-based

data to permit search or query capability and retrieval.

Conventional models for document imaging systems involve usage of

imaging equipment and software at a single central facility, e.g., at a hospital or

bank, under the control and direction of a central computer at that facility. Under this model, however, companies having multiple offices, desiring to centralize their records, have to ship their documents (either physically or electronically) to a central computer for centrally storing all of the documents

5   and permitting access via phone or other dedicated lines.

Despite the advent of networking, e.g., local area networks or LANs and now the Internet, this central computer model has nonetheless retained hold. With the emergence of the Internet as a platform for commerce, however, new paradigms of operation became possible. Instead of companies investing

10   heavily in equipment and manpower to support the scanning, indexing and storage of their own documents, companies could eliminate this entire overhead by outsourcing these and other data management functions. Applicants have recognized the need for this and other such services and have designed an improved system and methodology for servicing this heretofore unrecognized

15   but greatly desired need.

It is, therefore, an object of the present invention to provide an improved system and methodology for document storage, management and retrieval.

It is also an object of the present invention to provide an improved remote distributed capture system, eliminating the need to ship documents to a central point for processing.

5 SUMMARY OF THE INVENTION

The present invention is directed to a document imaging platform system and methodology for capturing, transmitting, storing, retrieving and displaying documents in a shared-system environment using the Internet or other network. Through utilization of thumbnail images along with full images, transmission

10 of multipage documents is facilitated, avoiding system bottlenecks. Document security is hierarchically based with document control being available to system users in addition to system administrators.

BRIEF DESCRIPTION OF THE DRAWINGS

15 The disclosed invention will be described with reference to the accompanying drawings, which show important sample embodiments of the invention and which are incorporated in the specification hereof by reference, wherein:

FIGURE 1 is a diagram of a system incorporating the principles of the present invention; and

FIGURES 2A and 2B are diagrams illustrating functionality configurations pursuant to the teachings of the present invention.

5

DETAILED DESCRIPTION OF THE PRESENTLY PREFERRED EXEMPLARY EMBODIMENTS

The numerous innovative teachings of the present application will be

10 described with particular reference to the presently preferred exemplary embodiments. However, it should be understood that this class of embodiments provides only a few examples of the many advantageous uses of the innovative teachings herein. In general, statements made in the specification of the present application do not necessarily delimit any of the various claimed inventions.

15 Moreover, some statements may apply to some inventive features but not to others.

With reference now to FIGURE 1 of the Drawings, there is illustrated an exemplary embodiment of a system configuration pursuant to the teachings of the present invention and generally referred to by the reference numeral 100. In

20 particular, there is illustrated a system and methodology for capturing document

images, transmitting these images to an image repository and permitting access to those stored images by an outside user.

At the user side, designated generally in FIGURE 1 by the reference numeral 101, a document (or more generally an object) may be captured, i.e.,

5   scanned by a scanner in a conventional manner. Each document image is, upon capture, assigned a unique system-wide image identification number, with this image identification number being stored in a central database repository 134. The database repository 134 stores information about each image, e.g., the image location (whether remote or local) and whether the image needs to be

10   transmitted. The images are stored in a central image repository 132, accessible via a repository interface 130. Further details about image storing are discussed further hereinafter.

During capture, designated generally by the reference numeral 102, the document image is temporarily stored in a local cache directory 118 for

15   subsequent transfer. It should, of course, be understood that after capture, the document may be indexed, providing a variety of textual indicators useful in later identifying that particular document from a potential myriad of similar-looking documents.

Indexes applied to and associated with each image are stored in the central database repository 134, and are referenced by the image identification number assigned during scanning. With further reference to FIGURE 1, the indexing can be facilitated through use of a Configurator 140, providing

5   additional speed and quality enhancements to processing, including cleanup of images (designated generally by the reference numeral 146), extraction of barcode values to automate data entry (designated generally by the reference numeral 148), and using database lookups to automatically populate index values. Each of these configuration parameters is preferably available to the

10   user through a graphical user interface, requiring no specialized computer programming skills to implement.

With reference again to FIGURE 1, a cache controller 120 communicates with the central database repository 134 through a cache server 122, e.g., a Web server (generally designated by the reference numeral 124), to

15   determine whether there are any images in the local cache, i.e., the cache adjacent capture 102 (cache 118), that need to be transmitted to the central image repository 132. For each image that needs to be transmitted, the cache controller 120 sends the image to the central image repository 132 via the Web server 124, and the image record in the central database repository 134 is

correspondingly updated with the image location. When indexing is complete

at the user side, and all indexed images have been transmitted for storage, the

cache controller 120 is notified by the database repository 134 that the

document images in the temporary cache, e.g., cache 118,may be deleted.

5   It should, of course, be understood that the Web server 124 used for

communications between the user side and the server side should be configured

for maximum encryption or other security algorithms to maintain data privacy,

and hinder eavesdropping or other potential intrusions.

   It should be understood that the respective capture stations 102 are

10 assigned a unique system-wide identification number, which allows the central

servers to reliably know where data is coming from. Once a batch has been

created, the central site, i.e., the repository, maintains an audit record of

everything that happens to the batch throughout its life-cycle, and this audit is

available real-time to the user. Transfer of batch data happens real-time during

15 scan and index. Transmission of images is offloaded to an unattended

application, the cache controller 120. The cache controller 120 receives all

instructions from the central server, which tells it which images need to be

transferred and which batches are eligible for deletion from the remote cache.

The central server also provides an operations person the ability to schedule

when each individual remote site 118 can send images (the transmission window), allowing Applicants to level-load the network bandwidth.

It should be readily understood that additional devices may be employed to forward non-indexed images to the central image repository 132 for storage

5   therein. For example and with reference again to FIGURE 1, a facsimile machine 110 maybe used to forward a document. For example, the facsimile machine 110 could forward the image to a facsimile server device 126, and from the facsimile server device 126 to an electronic mail server 128 for transmission, via an email import (designated generally by the reference

10   numeral 129), to the repository interface 130 to the central image repository 132.

In a similar fashion, a networked digital scan device 112, such as a Digital Sender device made by Hewlett Packard, or a Document Centre device 114, such as made by Xerox, may be employed to scan and forward document

15   images as electronic mail attachments to the electronic mail server 128, as discussed hereinabove. Additionally, an electronic mail application (designated generally by the reference numeral 116) could be used to forward document images as attachments to the electronic mail server 128. As discussed, all of the transmissions to the e-mail server 128 are forwarded to the image repository

132 by the e-mail import device 129. It should be understood that the e-mail import device 129 also created image records in the database repository 134, and identifies these image records as non-indexed. The non-indexed records are then available for indexing by any capture station 102 having access to the

5   central system.

Whereas the above describes various mechanisms for the capture and transference of images and index records to the central repository, additional functions permit management and manipulation of the images on the repository side.

10   With reference again to FIGURE 1, the various functions performed by the system, whether at the client side or repository side, are tracked and logged to a system journal 142. It should be understood that the system journal 142 creates records for such items as security violations, creation and transmission of images and indexes, metrics involving the time it

15   takes to perform actions and who performs them, as well as audit records for who has accessed the system and what documents they have viewed.

Contents of the system journal 142 are available through an Audit and Reports interface (designated generally by the reference numeral 108) and

provide the user with ad hoc report generation capabilities on any system activity.

With further reference to FIGURE 1, particularly regarding repository-side functions, image index records are preferably aggregated into document

5 records, each document record being a unique instance of all of the contextual index values applied to individual images. This process of creating document records is performed by a document maker 144, which builds the document records and preferably creates a thumbnail representation of each image within the document.

10 In addition to the storage of full images of documents within the repository, smaller versions thereof, i.e., thumbnail images, are also stored. Upon document capture 102 and creation, the respective images and indices corresponding thereto are transmitted to the repository, where the respective thumbnail images are created and stored in a single file by system identification

15 number. All document images (pages) are preferably stored in a single image format (no multipage TIFFs). When a user selects a document for viewing, the first page of the document is sent in full along with preferably all thumbnail images for all pages. Since the thumbnail images are considerably smaller than the original or full image size, e.g., the thumbnail image being less than about

one kilobyte in size, the user has the ability to see a representation of all of the pages of the document in order to make a further viewing selection.

It should be readily apparent to one skilled in the art that minimizing data transference in this fashion greatly improves system performance and

5   avoids unnecessary downloads.

Other repository-side functions are available to manipulate the images, including image processing techniques and other image cleanup techniques, described hereinabove in connection with reference numeral 146. Barcode processing can also be performed at the repository side to automatically extract

10   index values from one or more barcodes affixed to the document images, also described hereinabove in connection with reference numeral 148.

Documents stored on the system are, of course, made available for search and display by a user. For example, and with reference again to the system configuration illustrated in FIGURE 1, logging onto a client-side

15   computer, generally represented by the reference numeral 104, interfaces the user with a Web server 124, which provides access to the central database repository 134 and the central image repository 132 via the repository interface 130. This logon validates the user and the sections of the repository, or subset of documents, that the user has access to. Similarly, a query may be made at

the client-side computer 104 to search for various documents stored within the repository using a number of search indices. Client execution of a query causes the repository interface 130 to generate a list of documents matching the search criteria, and returns the list to the client computer 104. The user may then

5   select a given document from the list for viewing, the selection causing the first page of a multiple page document and thumbnail images of all of the remaining document images to be retrieved from the image repository 132 via the Web server 124 and displayed on the client computer 104. Full images of any subsequent pages of the multipage document are retrieved from the image

10   repository 132 only if requested by the user, e.g., by clicking on a thumbnail image displayed to the user,

The advantages of this system configuration over prior document centralization methods are manifest. Instead of shipping countless pounds or tons of documents to a centralized facility for scanning, indexing and storage,

15   this functionality can be distributed with scanning and perhaps indexing being localized endeavors, e.g., at bank branches, and electronic storage being centralized at the headquarters of an organization. Outsourcing one or more of these functions, expensive equipment, know-how and manpower can save a company considerable sums and generate efficiencies within the organization

12093

by simplifying the processing of and access to such data. Further, the cost of

riding the technological wave of new software and hardware, perhaps the

bleeding edge thereof, can be avoided through such outsourcing, leaving the

responsibility for technological advancement and capability in the hands of

5   those skilled in the area.

With reference now to FIGURE 2A, there is illustrated a functional

overview of the system configuration according to the present invention,

designated generally by the reference numeral 200. Applicants have created

various software tools to facilitate user interaction with the data stored in the

10   repository. An Application Program Interface 202, for example, facilitates the

aforedescribed capture 102 and cache controller 120 functions, designated

generally by the reference numerals 204 and 206, respectively, along with a

variety of vertical applications 208. The API 202 also governs communications

using ActiveX commands, e.g., an ActiveX Query 210 and an ActiveX Viewer

15   212, both in communication externally via a portal integration node 214. Java

Queries 216 and Java Viewers 218 communicate with an Application Server

220.

Both the API 202 and the Application Server 220 govern contact with a

backend program 222, e.g., the aforedescribed repository interface 130 in

FIGURE 1, which controls communications with database services 224 and

image services 226, e.g., the aforedescribed database repository 134 and image

repository 132, respectively. As illustrated in FIGURE 2A, the database

services 224 governs configuration 228, indexing 230 and auditing 232, and the

5   image services 226 governs image storage 234, optical archiving 236, image

cleanup 238, data extraction 240 and image redelivery 242.

      With further reference to FIGURE 2B, there is illustrated a preferred

function configuration, designated generally by the reference numeral 250. A

cache controller 252 and a capture node 254 interface with an API 256, which,

10   in turn, communicates with a web server 258, e.g., a Microsoft Transaction

Server (MTS). Alternatively, a Java Viewer 260 may interface with an

application server 262. Both the web server 258 and the application server 262

communicate, via a backend program 264, to a database services 264 and an

image services 266, as discussed in more detail hereinabove in connection with

15   FIGURE 2A.

      In addition to offering an improved paradigm over conventional

document retention schemes, the present invention is also directed to

improvements in the accessing of such documents, offering new techniques in

security. As is understood in the art, security issues in the single facility model

are governed by an administrator who directly controls the administration of the entire system.

The present invention employs the hierarchical concept of an account, a domain, an application and index fields to categorize the information. For example, an account represents a contract with a customer, e.g., a university, to provide document imaging services. A domain is a facet of the account, e.g., a department within the university such as student records, and an application would represent instances of the domain, e.g., admissions or transcript records. The final layer of granularity is the index field, which defines documents in applications, e.g., student name. Instead of the system administrator for the account controlling access at all levels, i.e., no granularity of control, control or access can be granted to domains or applications, distributing security to end users in multiple tiers. In other words, the system and methodology of the present invention places full control of the lookup configuration directly in the user's hands and requires no special programming to implement.

An advantage of this approach is ready reconfigurability by the user instead of an administrator. A form of distributed security is possible where only viable index fields permissible to that user are presented and others masked. One mechanism for employing this aspect of the present invention is

having the user themselves use Open Database Connectivity (OBDC) protocols to define the index fields or lookups. By using a standard interface accessible to a variety of database formats, such as ODBC, the user instead of a system administrator can control or configure what they see and how.

5   It should further be understood that although the present invention is currently implemented in Visual BASIC with ActiveX controls, additional software tools may be employed to practice the principles of the present invention. For example, at least one such software tool is JAVA, which would offer additional benefits to this innovation.

10   As will be recognized by those skilled in the art, the innovative concepts described in the present application can be modified and varied over a wide range of applications. Accordingly, the scope of patented subject matter should not be limited to any of the specific exemplary teachings discussed, but is instead defined by the following claims.

12093